The National Library of Medicine Pill Image Recognition Challenge: An Initial Report

Ziv Yaniv[1,2], Jessica Faruque[1,3], Sally Howe[1], Kathel Dunn[1], David Sharlip[1], Andrew Bond[4], Pablo Perillan[4], Olivier Bodenreider[1], Michael J. Ackerman[1,2], Terry S. Yoo[1]

[1]National Library of Medicine, National Institutes of Health, Bethesda, MD, USA
[2]TAJ Technologies Inc., Mendota Heights, MN, USA
[3]CytoVale Inc., San Francisco, CA, USA
[4]Medicos Consultants LLC, Baltimore, MD, USA

Abstract—In January 2016 the U.S. National Library of Medicine announced a challenge competition calling for the development and discovery of high-quality algorithms and software that rank how well consumer images of prescription pills match reference images of pills in its authoritative RxIMAGE collection. This challenge was motivated by the need to easily identify unknown prescription pills both by healthcare personnel and the general public. Potential benefits of this capability include confirmation of the pill in settings where the documentation and medication have been separated, such as in a disaster or emergency; and confirmation of a pill when the prescribed medication changes from brand to generic, or for any other reason the shape and color of the pill change. The data for the competition consisted of two types of images, high quality macro photographs, reference images, and consumer quality photographs of the quality we expect users of a proposed application to acquire. A training dataset consisting of 2000 reference images and 5000 corresponding consumer quality images acquired from 1000 pills was provided to challenge participants. A second dataset acquired from 1000 pills with similar distributions of shape and color was reserved as a segregated testing set. Challenge submissions were required to produce a ranking of the reference images, given a consumer quality image as input. Determination of the winning teams was done using the mean average precision quality metric, with the three winners obtaining mean average precision scores of 0.27, 0.09, and 0.08. In the retrieval results, the correct image was amongst the top five ranked images 43%, 12%, and 11% of the time, out of 5000 query/consumer images. This is an initial promising step towards development of an NLM software system and application-programming interface facilitating pill identification. The training dataset will continue to be freely available online at: http://pir.nlm.nih.gov/challenge/submission.html

.

Keywords—prescription pill images, content-based image retrieval, image matching, open data

I. INTRODUCTION

Correct identification of prescription oral solid dosage forms of medication, pills, based on their visual appearance is a key step required to assure patient safety and facilitate more effective patient care.

From a safety perspective, a variety of errors can occur across the pharmacological chain. These include errors in prescription, transcription (interpretation of medication order), dispensing, and administration [1]. Such errors can result in adverse drug events (i.e. injuries resulting from medication use). Identification and confirmation of prescribed pills can mitigate dispensing errors. While dispensing error rates are low, pharmacies continue to incorporate technologies (e.g. bar code systems) to further reduce them [2]. Visual identification is an additional technology addressing this issue, which can also be deployed outside the pharmacy for use prior to medication administration. This technology is potentially usable by both healthcare personnel and the general public.

For patient care, identification of pills based on their visual appearance has the potential to facilitate improved treatment in a variety of settings, including disasters, poison control centers and patient persistence.

Following natural or man-made disasters, it is not uncommon for pills and documentation to be separated. A recent literature survey identified medication loss and difficulties in filling prescriptions as common phenomena in evacuation scenarios, with evacuees only able to provide partial information about their medications [3]. This is of particular concern for people with chronic conditions that may experience adverse effects with a disruption in their medication regimen. The need to fill in missing information and identify medications is also illustrated in the operational flow of a successful evacuation center operated during the hurricane Katrina event [4]. Part of the center's medical screening examination included medication identification by a pharmacist.

Another aspect of patient treatment in which visual identification of pills can potentially improve care has to do with the availability of experts as part of the poison control centers system. U.S. poison control centers provide expert advice, over the telephone, with respect to possible exposure to poisonous substances, such as some prescription medications. The service is available to healthcare personnel and the general public. Analysis of call volumes to poison control centers has shown a significant increase, primarily associated with requests for pill identification [5], [6]. This has proven to be a burden on the centers as it occupies the limited number of experts, diverting their attention from more critical calls. One effective solution to this issue was described in [7]. Using an interactive

voice response system to provide automated pill identification reduced the call volume requiring a human expert by more than 50%, allowing the center's experts to address the more critical calls.

Finally, identification of pills has the potential to increase patient persistence, "the duration of time from initiation to discontinuation of therapy" [8], when the visual appearance of a prescribed pill changes[1]. This is of particular importance to patients dealing with chronic conditions such as hypertension or epilepsy as they are likely to experience these changes multiple times. The common reasons for change in color or shape is switching between equivalent brand-name and generic pills and between generics. Financial incentives can also drive frequent changes between equivalent generic versions of the same drug. In several studies it has been shown that switching between brand-name and generic versions that are visually different resulted in reduced persistence [10], [11]. Other studies have found this not to be the case, with the change having no effect [12], [13]. This conclusion was qualified by the fact that these results were obtained when the change was physician initiated and focused on the initial change from brand-name to generic. Finally, some studies were less conclusive with part of the population exhibiting increased or similar persistence and some decreased persistence [14], [15]. Increased persistence was attributed to lower costs associated with generics and to past experience switching between products. Reduced persistence was associated with patients experiencing their first change. Pill identification can reassure patients that they are taking the appropriate medication, potentially increasing their persistence.

In this work we describe the design and results of the National Library of Medicine (NLM) challenge for development of algorithms and software that rank reference images in the authoritative NLM RxIMAGE collection with respect to their visual similarity to given consumer quality query images. In this context, consumer quality images are pictures of a single pill placed on a uniform background without any constraint on the camera pose (i.e. images do not have to be acquired perpendicular to the pill face).

It should be noted that the challenge's goal was not to identify pills based on their visual appearance. This decision was motivated by two observations. First, identification would have placed unnecessary requirements on an algorithm's performance as compared to retrieval of the correct reference image amongst the first n retrieved images (specific values for n are task dependent). The person submitting the query is then charged with identifying the pill by selecting the image and accompanying information from the retrieved ones. Second, identification of pills by their visual characteristics is all but impossible when considering the existence of counterfeit products, many of which are visually indistinguishable from the real product [16]. In this case, identification of the medication

is impossible as similarity to the relevant reference image does not imply the product contains the expected ingredients.

Currently, there are a variety of commercial products and web based services for pill identification. None of these provides a complete solution that is readily available and easy to use for both healthcare professionals and the general public. NLM currently provides two such web based services, DailyMed [17] and Pillbox [18]. DailyMed is primarily intended for use by healthcare professionals. This service displays manufacturer-provided labeling information that is retrieved based on the product's National Drug Code (NDC) or keywords that appear in various sections of the label. Pillbox is more appropriate for the general public as it displays information based on textual input such as the pill's shape or color. A large number of web sites offer similar functionality to that provided by Pillbox, including among others WebMD [19], RxList [20], and GoodRx [21]. The main disadvantage of all of these sites is that input is textual and it is up to the user to select the right pill characteristics, which remains challenging (e.g. which color should one select with a two color capsule?).

With the widespread availability of digital cameras, using a photograph of a pill is a more convenient form of input compared to textual input, an observation that both researchers and companies have made.

Several commercial products for pill identification based on photographs are available on the market. The MedEye system from Mint Solutions (Amsterdam, Netherlands) is designed for use in hospital settings, with several different pills inserted into an enclosed pod that acquires images under controlled lighting and uses these to identify the pills. These are then compared to the prescribed medications. A similar tabletop device is IdentRx from PerceptiMed (Mountain View, CA, USA). Similarly to the MedEye system, pills are placed into the device with images acquired under controlled conditions. The target audience includes hospitals, pharmacies, and long term care facilities. The MyPillSense system from IRODY (Boston, MA, USA) provides a custom smartphone application that, according to the company, is combined with computer vision and AI algorithms to identify pills. The image acquisition application guides the user using an overlay so that the pills of interest, possibly several different products, are in a specific location in the video. The MedEye, IdentRx, and MyPillSense systems compare the identified pills to the patient's prescribed medication regimen. It is unclear whether the known medication regimen is used to limit the search space of possible pill types. The MedSnap system from MedSnap (Birminghan, AL, USA) also provides a smartphone application for pill identification. It uses an overlay to guide image acquisition in conjunction with a customized background surface that facilitates color and size measurements. The pill is identified without any usage of medication regimen information.

In the academic literature the challenge of pill identification has been addressed by several groups [22], [23], [24], [25], [26]. All of these approaches share a common theme, the use of feature vectors designed based on domain knowledge with ranking based on the distances between the feature

---

[1]Adherence is often used in the literature to refer to the period of treatment even though its definition refers to following the prescribed instructions which are a combination of dosage, timing, frequency and treatment period [9].

Fig. 1. Reference image macro photography rig. Inset shows the macro calibration card placed on top of floating stage and the imaged pills.

vectors of the reference images and consumer, query, images. In [22] the features encode shape, color and imprint. This work simulated the consumer quality images by translating, rotating and modifying the brightness of the reference images. Unfortunately, this does not truly represent the variability of consumer images. In [23] the features included the Scale Invariant Feature Transform (SIFT) descriptors, color, and multiscale local binary pattern. Pills were localized in the query image by cropping the image using the bounding box of non zero gradient magnitude. This approach is unlikely to be applicable for the less controlled imaging conditions that we envision. In [24] a customized background surface with grid lines is used to estimate pill sizes with features including color, shape, length to width ratio, size and texture. In [25] a customized background surface is used, again to enable the estimation of pill size. Additional features used include Hu moments as shape descriptors and color. As the aim of this work was to provide an assistive system for the elderly, the query images are compared to the small subset of products in a person's medication regimen and not necessarily to a general database of reference images. Finally, [26] use color and shape features with an emphasis on representing the shape of the imprint. The query images in this evaluation appear to be limited to those where the imprint is clearly visible and images are acquired with the camera perpendicular to the pill surface.

None of these pill identification systems provides the solution we set out to obtain in the challenge. Some use higher quality query images than usually obtained by simple point and shoot camera applications such as those that come pre-installed with mobile devices. Others require more controlled imaging conditions, controlled via customized applications possibly in conjunction with custom background surfaces.

II. MATERIALS AND METHODS

To run all entries on a single physical system, challenge submissions were required to provide their entries using a virtual machine (VM). The accepted VMs included several variants of Linux, OS X and Windows. The physical system we used was an Apple Mac Pro with the following configuration: 3.7GHz Intel Xeon E5 quad-core CPU, 512GB PCIe-based flash storage, 64GB 1866MHz DDR3 ECC RAM, OS X 10.10, VMware Fusion 7.0.

As the challenge required that we run unknown software whose origins are unverified, for security reasons, the evaluation system was not connected to any network. Consequentially, participants were required to physically mail their submissions on a USB drive.

A. Data

The challenge datasets consist of two types of photographs, high-quality macro photographs, reference images, and consumer-quality photographs of the quality we expect users of the proposed application to acquire.

1) Data Acquisition Process: Acquisition of reference images used a consistent workflow to ensure data integrity.

Products were acquired from multiple licensed sources including: (1) accredited retail pharmacies with audited supplies of ethically marketed products in the United States; (2) directly shipped from pharmaceutical manufacturers in response to NIH federal register notice (FR79 FR 56381), and (3) from an accredited school of pharmacy. All products were managed through a Centralized Inventory Management System (CIMS) developed by Medicos Consultants, LLC (Baltimore, MD, USA). This system utilizes barcode identification of the product's NDC number provided by the pharmaceutical manufacturer or relabeler. Beyond the NDC barcode information, the CIMS system maintains additional specific product information for every pill, including: product source, labeler name, product name, dose strength, one-of-multiple, lot number, expiry date, storage location, variance between pill samples, workflow process ticketing information, and records of qualified individuals handling each workflow step. Every physical sample is assigned a unique inventory container which is tracked using multiple validation points throughout the the workflow which includes the following steps: acquisition, inventory management, imaging, quality control, image processing, and final image and product review. An integrated ticketing system allows for the interaction of multiple team members to reference particular pill samples and processing dates in an auditable and traceable manner.

Integrating with the CIMS technology framework are several components that are critical to maintaining the integrity of the product supply, all of which were developed by Medicos Consultants LLC. These include PharmFinder®, PharmImage®, PharmImageProcess®, and PharmDataDistribution®.

PharmFinder is a mobile system that allows trained pharmacists and pharmacy technicians to remotely scan in the UPC barcode on a candidate product labeling to determine if images of that product have already been acquired. If the product is not part of the image collection, the application flags it for dispensing. Pharmacists and pharmacy technicians

Fig. 2. Sample images from the training dataset. The first two columns are the reference, macro photographs. The remaining pictures are consumer quality images.

review the NDC and physical form to ensure that the product is relevant for inclusion in the database. The application is also used to enter in notifications of shipments of products directly sent from manufacturers so that they can be tracked from the moment we are notified of the the shipment to arrival. Using PharmFinder in conjunction with the CIMS we validate the product name (trade/generic), NDC, product dosing, manufacturer, labeler and physical characteristics.

PharmImage is an image acquisition suite that controls and monitors the specialized macro photography rig as well as multiple camera configurations, including: camera controls, lens parameters, illumination adjustments and related imaging conditions.

PharmImageProcess is a distributed image processing suite. Segmentation and final image compositing begins with automated macro calibration card processing, starting with recognition of the four 2D fiducial registration marks, collection of incident and white balance illumination data for the specimen. Image processing using the CIMS and PharmImageProcess allows segmentation of the individual specimens to occur as part of an integrated workflow. Constraining the entire lifecycle within the CIMS eliminates product association errors that otherwise may occur if individual image files were manually processed using multiple applications.

Finally, the PharmDataDistribution component manages the secure distribution of project data to various remote locations including those at multiple imaging laboratories, remote pharmacies, image processing locations, as well as the distribution of the final data to the NLM.

To obtain high quality macro photographs we used the acquisition rig shown in Fig. 1. The camera utilized was a Canon EOS 5D Mark II (21.1 MP) with a 100mm macro lens. The rig was controlled by the PharmImage software. Both non-polarized and cross polarized illumination was used. Cross polarization is used to reduce specular highlights, glare and reflections from shiny surfaces such as capsules. Color, white balance, size and rotation calibrations were performed using a macro calibration card, see Fig. 1 inset. Finally, to facilitate segmentation the pill was positioned on a "floating" stage above a transparent, neutral Snell-Descartes (incidence) background.

Consumer-quality (query) images were acquired by placing a single pill on a uniform background with no restrictions on camera pose. A variety of mobile phone and consumer grade cameras were used to acquire the images. To simulate varying illumination, we used the following light sources: fluorescent, incandescent, LED, CFG as well as direct and indirect diffusion filtering.

2) Data characterization: The data used in this work had three components: a training dataset provided to all challenge participants, a segregated testing dataset and a non-segregated testing dataset. The non-segregated testing dataset was only used to evaluate the three top submissions after the challenge concluded.

The training dataset consists of 7000 images corresponding to 1000 pills. For each pill there is one high quality macro photograph of each side of the pill and five consumer quality images. For capsules the definition of side was based on the imprint being fully visible with a perpendicular view of the text "plane". The pills selected for this dataset were randomly selected from the RxIMAGE collection so that their distribution based on color and shape matches that of the complete dataset. Examples of reference and consumer quality images for a tablet and capsule are shown in Fig. 2.

The segregated testing dataset was created using an additional set of 1000 pills that were not seen by the participants. This dataset consisted of two reference images and five consumer images per pill. Pills were also randomly selected using their shape and color such that their distribution was similar to that of the training set. A bar chart showing the pill distributions based on their color and shape is shown in Fig. 3. This dataset allows us to evaluate the generalization ability of an algorithm and best fits our long-term goals, as we continue to acquire additional images for the RxIMAGE collection and prefer to train an algorithm only once.

Finally the non-segregated testing dataset consisted of data from the same 1000 pills used in the training set. In this dataset we used 6486 consumer quality images that were not part of the training set. This data allows us to evaluate whether an algorithm is possibly over-fitting to the training data, as it should generalize well to variations of consumer images similar to those it has previously trained on.

B. Evaluation Metrics

Evaluation of all submissions was done using a black box evaluation approach. That is, we considered only the input
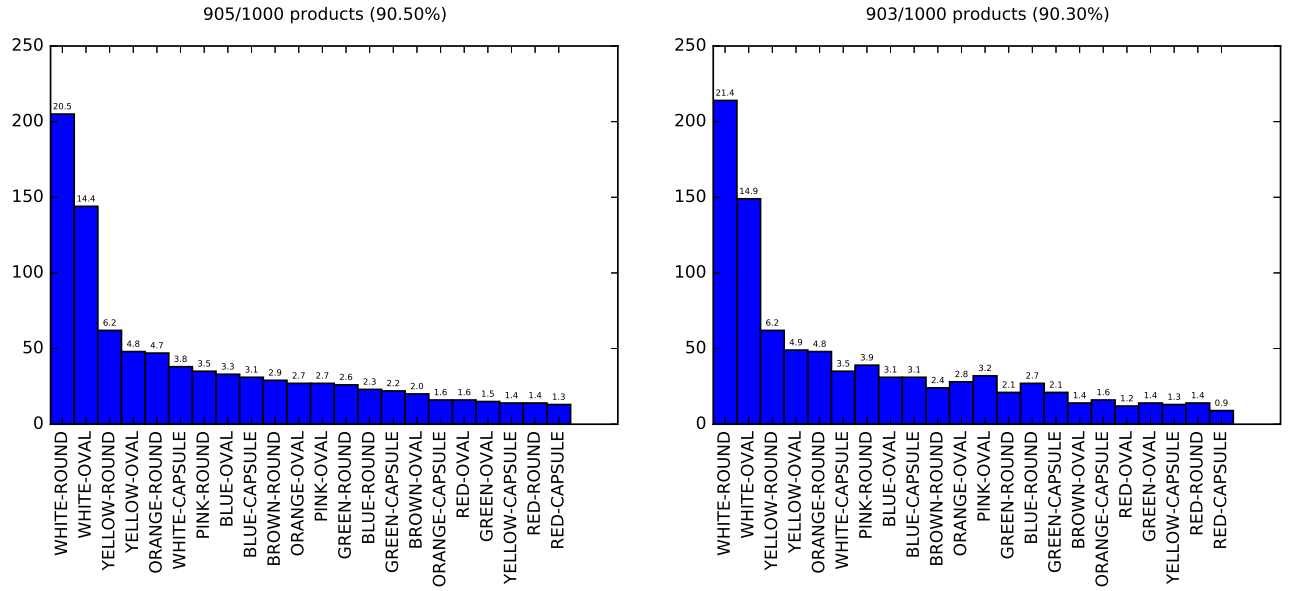
905/1000 products (90.50%)

20.5 14.4 6.2 4.8 4.7 3.8 3.5 3.3 3.1 2.9 2.7 2.7 2.6 2.3 2.2 2.0 1.6 1.6 1.5 1.4 1.4 1.3

WHITE-ROUND, WHITE-OVAL, YELLOW-ROUND, YELLOW-OVAL, ORANGE-ROUND, WHITE-CAPSULE, PINK-ROUND, BLUE-OVAL, BLUE-CAPSULE, BROWN-ROUND, ORANGE-OVAL, PINK-ROUND, GREEN-ROUND, BLUE-ROUND, GREEN-CAPSULE, BROWN-OVAL, ORANGE-CAPSULE, RED-OVAL, GREEN-OVAL, YELLOW-CAPSULE, RED-ROUND, RED-CAPSULE

903/1000 products (90.30%)

21.4 14.9 6.2 4.9 4.8 3.5 3.9 3.1 3.1 2.4 2.8 3.2 2.1 2.7 2.1 1.4 1.6 1.2 1.4 1.3 1.4 0.9

WHITE-ROUND, WHITE-OVAL, YELLOW-ROUND, YELLOW-OVAL, ORANGE-ROUND, WHITE-CAPSULE, PINK-ROUND, BLUE-OVAL, BLUE-CAPSULE, BROWN-ROUND, ORANGE-OVAL, PINK-ROUND, GREEN-ROUND, BLUE-ROUND, GREEN-CAPSULE, BROWN-OVAL, ORANGE-CAPSULE, RED-OVAL, GREEN-OVAL, YELLOW-CAPSULE, RED-ROUND, RED-CAPSULE

Fig. 3. Distribution of pills based on their key visual properties, shape and color. (left) training data and (right) segregated testing. The segregated testing dataset is comprised of reference and consumer quality images from products that were not part of the training set.

and output of the programs without any consideration of the inner workings of each of the submissions. We used common information retrieval metrics [27].

In the challenge, the only criterion on which submissions were assessed was the Mean Average Precision (MAP):

$$MAP = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{N_i} \sum_{j=1}^{N_i} \frac{j}{MT(i,j)} \right]$$

where N is the number of consumer-quality images, N_i is the number of matches for consumer quality image i, j is the number of correct matches, in our case one or two, and MT is the rank of the correct reference images, $MT(i,j) < MT(i,j+1)$. In our evaluation we used the ranks of both reference images, both sides of the pill, because images of capsules often do not have a "correct" side.

For the three top submissions we also looked at the Mean Reciprocal Rank:

$$MRR = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{MT(i)}$$

where $MT(i)$ is the rank of the first correct reference image.

We also recorded their runtime after modifying the submitted VMs so that they all had the same computing resources. Namely, we allocated each of the VMs 4 cores and 16Gb of RAM. On Linux we used the time command and on Windows the %time% command to get the wall-clock time before and after program execution.

C. Challenge Submissions

We received 11 challenge submissions. Four of these utilized deep learning approaches [28] with the remaining en-

TABLE I
MAP (MRR) SCORES FOR THE TOP THREE SUBMISSIONS.

Dataset / Submission	Training	Segregated Testing	Non-segregated Testing
msumpf	0.41 (0.54)	0.27 (0.34)	0.28 (0.35)
castelo	0.12 (0.13)	0.09 (0.09)	0.08 (0.08)
nhatuntsev	0.12 (0.13)	0.08 (0.08)	0.10 (0.09)

tries using the methodology previously applied to this task, designing feature vectors based on domain knowledge.

The three top teams included one entry using the traditional approach, team "nhatuntsev", with distances in feature space defined as a weighted sum of distances in color space and shape space. The two other teams used deep learning approaches. Team "castelo" used a Convolutional Neural Network (CNN) implemented with Google's TensorFlow open source library. Team "msumpf" used a combination of features obtained using deep learning and the SIFT descriptor. This entry used a CNN implemented with University of California Berkeley's Caffe open source framework; image similarity was defined as the weighted sum of the two similarity scores with the weights being 0.8 and 0.2 for CNN and SIFT, respectively.

We next provide the detailed results analysis for the top three teams.

III. RESULTS

The official challenge outcome was based on the MAP score obtained by the submissions on the segregated testing data, with the MAP scores for the top three teams being: "msumpf" – 0.27, "castelo" – 0.09, and "nhatuntsev" – 0.08.
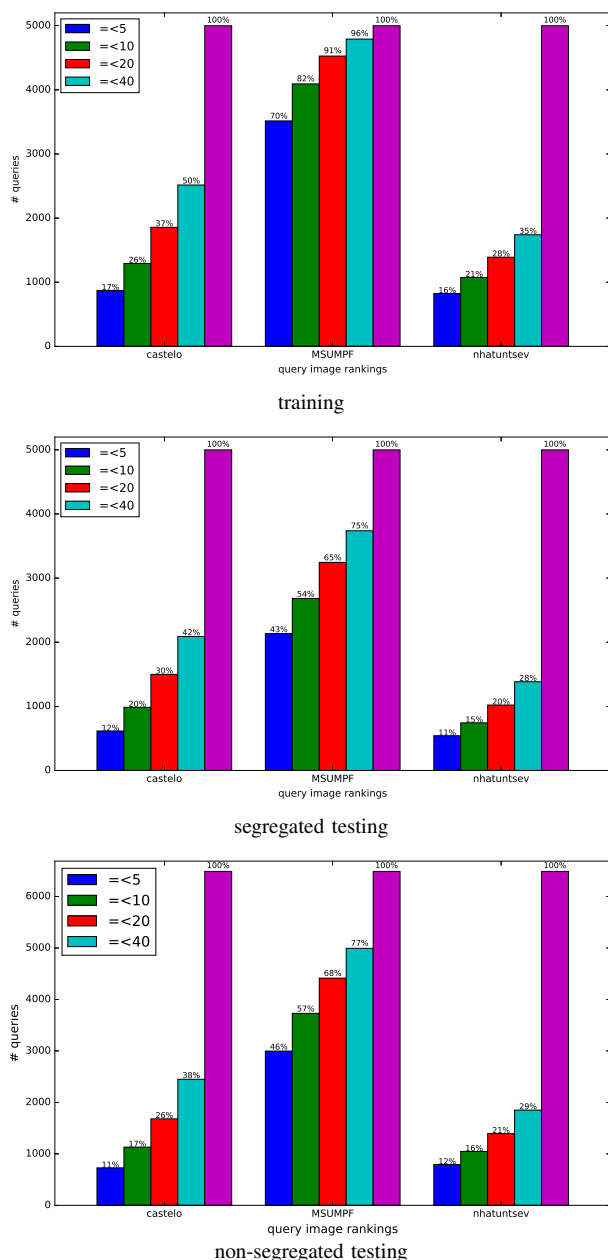
Fig. 4. Number of queries and percentages of correct reference images in the first 5, 10, 20 and 40 retrieved results. All approaches do the best on the training set, as the algorithms are "optimized" on this data. Interestingly, the performance difference between segregated testing data and non-segregated testing data was small.

Further evaluation across all three datasets showed that the approach taken by team "msumpf" yielded the best results, though it appears to exhibit performance associated with overfitting, showing very good performance on the training dataset and degrading considerably on the two testing sets. Table I summarizes the results obtained by all three top teams.

As MAP and MRR only provide a summary of a submission's average performance on all queries we also evaluated each submission's performance using the distribution of ref-

erence image rankings given each query image. In our case we looked at the percentage of queries in which the correct reference image was in the first 5, 10, 20, and 40 images. These specific numbers are based on our expectation that results will be displayed on a single page. With a mobile device we expect 5-10 results would fit on screen, while with a standard screen we expect to display 20-40 results on a single page. Fig. 4 summarizes these results. An interesting observation is that all three submissions appear to have consistent performance on the segregated and non-segregated testing datasets. The expectation was that the performance on the non-segregated dataset would be considerably better, as the data originates from the same pills as in the training set and the new consumer quality images are similar to the ones in that set.

We next analyzed the results of the top-performing algorithm, team "msumpf", using a color-shape stratification of the segregated testing data. For the two most common strata, white-round and white-oval, the algorithm's performance degraded considerably. Respective MAP (MRR) scores were 0.08 (0.13) and 0.09 (0.14) compared to the overall score on the whole dataset 0.27 (0.34). The remaining strata had considerably better performance as illustrated in Fig. 5. Unfortunately, these two strata cannot be ignored as they account for 36.3% of the pills.

We then analyzed the performance of team "msumpf" using text imprint type as the stratification criterion, text is either printed or embossed/debossed. It should be noted that all capsules have printed text. We therefore refined the stratification to separate capsules from other printed pills. The testing dataset consists of 165 printed capsules, 62 printed tablets and 773 embossed or debossed pills. The results in this case were clear-cut, with capsules readily identified with 94% of the queries returning a correct match in the first 40 retrieved images. Similar performance is observed for printed tablets with 90% of the queries returning a correct match in the first 40 queries. When compared to the 69% success rate for the embossed/debossed imprints, it is clear that printed text plays a key role in facilitating identification. The respective MAP (MRR) scores for capsules, printed tablets and embossed/debossed tablets were 0.49 (0.63), 0.35 (0.59), and 0.21 (0.26). Fig. 6 summarizes this evaluation.

Finally, while optimal runtime of the programs was not our goal in this challenge, we did see significant differences between runtimes on the testing dataset with total runtimes of 44min ("nhatuntsev"), 1967min ("castelo"), and 84min ("msumpf").

IV. DISCUSSION AND CONCLUSIONS

The goal of the NLM pill image recognition challenge was to encourage the development and discovery of high-quality algorithms and software that rank how well consumer quality images of prescription pills, acquired with mobile devices, match high-quality macro photographs of them, reference images. This challenge was motivated by the need to easily identify unknown prescription pills both by healthcare personnel and the general public.
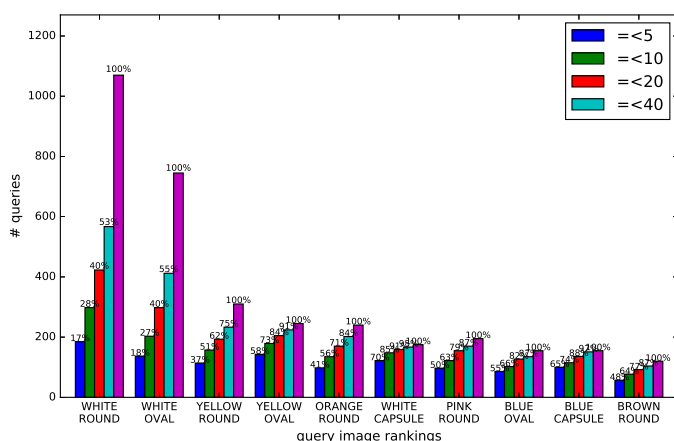
Fig. 5. Performance of team "msumpf" on the segregated testing dataset according to color-shape strata. Number of queries and percentages of correct reference images in the first 5, 10, 20 and 40 retrieved results.
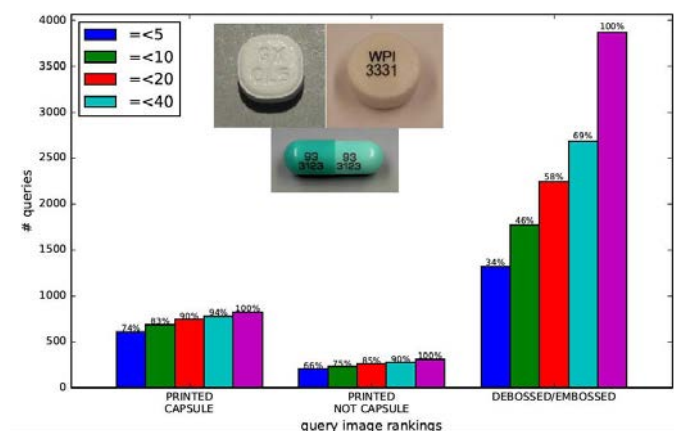

Fig. 6. Performance of team "msumpf" on the segregated testing dataset according to imprint type, printed (capsule/not capsule) or embossed/debossed. Number of queries and percentages of correct reference images in the first 5, 10, 20 and 40 retrieved results. Inset shows an example of an "easy" query image from each of these categories.

The reference images used in the challenge are part of the RxIMAGE collection, an authoritative image database of prescription oral solid dosage forms of medication, pills. The current collection contains high-quality macro photographs of more than 4000 pills that represent more than 40% of prescription pills (as a proportion of the NDCs). Images similar to the reference images, but containing additional information, are freely available from the collection via a RESTful application programming interface [29]. Additional retrievable information associated with each image includes product name, labeler, and the product's NDC. An example of the images available from the RxIMAGE collection is shown in Fig. 7.

The consumer-quality images used in the challenge were acquired with mobile devices using pre-installed camera applications. Each is a photograph of a single pill placed on a uniform background without any constraint on the camera pose. This represents the simplest image acquisition scenario, as the constraints on the user are minimal.

Analysis of the results obtained by the winning submission indicate that the type of imprint on the pill has a significant effect on identification. Capsules were more readily identified, with the correct reference image returned in the first five retrieved images 74% of the time. In this case textual information is printed and the pill colors are rather distinctive. Identifying tablets with printed text was slightly more challenging with the correct reference image returned in the first five retrieved images 66% of the time as compared to embossed/debossed pills where this happened only 34% of the time.

The performance exhibited by the top three submissions to the challenge is currently not sufficient for development of a mobile online service for matching reference images to consumer quality images similar to those used in this work. This is based on the expectation that we can display up to ten images on a mobile device's screen, while the winning submission returned the correct reference image in the first ten retrieved images only 54% of the time. On the other hand, in a healthcare setting with a professional viewing the results on a standard computer screen, the performance of the winning submission is possibly sufficient, with the correct reference image retrieved in the first 40 images 75% of the time.

The level of difficulty of matching images acquired with a mobile device to a set of reference images varies with context and the constraints placed on the acquired images. In the most generic context, the number of reference images is large, larger than the current RxIMAGE collection. This is the situation encountered in a poison control setting when the pill in question can be any pill on the market, including both prescription and over the counter medications. In more controlled contexts, such as applications to ensure safe and persistent medication usage in the home or hospital setting,
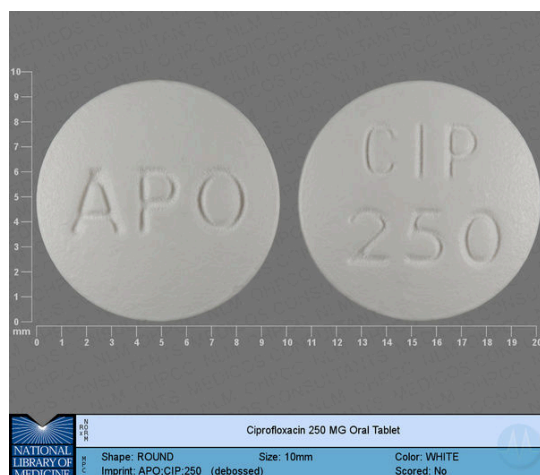

Fig. 7. Sample image retrieved from the RxIMAGE collection using the RESTful API.

the task is potentially easier. In this context, there is a much smaller number of reference images. The goal in such a setting is to confirm a pill's identity using the set of prescribed medications, indicating an issue if the pill image does not match the prescription list or any of the equivalent medications. The association between equivalent pills, generics and brand name, are readily available using the NLM's RxNorm information system. RxNorm provides a normalized naming system and tool for semantic interoperation between drug terminologies and is freely available online with a RESTful API [30].

The constraints one places on the content and acquisition of the query image have a significant effect on the complexity of the retrieval task. The least constrained and most challenging approach accepts query images containing multiple pills with any background, acquired with no constraint on camera pose. In this challenge we have chosen to place some limitations on the query images, namely that these were required to contain a single pill on a uniform background. More constrained settings have the potential to improve retrieval results. These can include placing objects of known shape, size, and color imaged alongside the pill. This can facilitate estimation of camera pose, pill color, and size. Actively controlling camera pose and the location of the pill(s) in the image is another approach. One can require the use of a customized acquisition application that overlays visual aides onto the live image, guiding the user to the desired camera pose. Finally, one can constrain the image acquisition via hardware with camera and lighting accurately positioned in known locations with respect to a tray containing the pills. All of these constraints have been used by existing solutions.

As our goal is to develop a generic application that is easily used in various contexts we placed minimal constraints on query image acquisition. Based on the challenge results it is possible that this initial approach to query image acquisition may have been too flexible. One possible path forward is to acquire consumer quality images with a known object, such as a coin, next to the imaged pill. This is a small modification to our expected usage of the system and does not require significant effort from the user.

Given the fact that the top two performing submissions utilized deep learning, their performance is affected by the size of the training dataset. In this case it is possible that the training dataset is not sufficiently large to achieve the desired performance. A common theme with deep learning algorithms is that they perform best when trained on a large set, hundreds of millions, of labeled examples [28]. In our case we provided 5000 consumer quality images (2000 of which were reference images), with only 2-3 consumer quality images per pill side. This likely necessitated the usage of techniques that mitigate the effects of a small training set in deep learning such as pre-training on an unrelated dataset and data augmentation, artificially generating additional samples via spatial and intensity transformations of the original images [31]. To better accommodate deep learning approaches we intend to investigate the use of cost-effective approaches for acquiring a large number of consumer quality images. One possibility for achieving this is to use a crowd-sourcing approach to data acquisition, often referred to as citizen science [32]. We will provide the set of training reference images and request the public to upload images of these products acquired with their mobile devices.

The results obtained by the algorithms submitted to the NLM pill image recognition challenge constitute an initial promising step towards development of a software system and application programming interface for pill identification. We expect to continue engaging the research community in future workshops, possibly a follow-on challenge, and by keeping the challenge training data freely available online from: http://pir.nlm.nih.gov/challenge/submission.html

REFERENCES

[1] P. van den Bemt and T. Egberts, "Drug related problems: definitions and classification," *EJHP Practice*, 2007.

[2] K.-C. Cheung, M. L. Bouvy, and P. A. G. M. De Smet, "Medication errors: the importance of safe dispensing," *British Journal of Clinical Pharmacology*, vol. 67, no. 6, pp. 676–680, 2009.

[3] S. Ochi, S. Hodgson, O. Landeg, L. Mayner, and V. Murray, "Disaster-driven evacuation and medication loss: a systematic literature review," *PLoS currents*, vol. 6, 2014.

[4] C. B. Irvin and J. G. Atas, "Management of evacuee surge from a disaster area: solutions to avoid non-emergent, emergency department visits," *Prehospital and Disaster Medicine*, vol. 22, no. 3, pp. 220–223, 2007.

[5] H. A. Spiller and J. R. K. Griffith, "Increasing burden of pill identification requests to US Poison Centers," *Clinical Toxicology*, vol. 47, no. 3, 2009.

[6] E. P. Krenzelok and R. Mrvos, "The use of an automated interactive voice response system to manage medication identification calls to a poison center," *Clinical Toxicology*, vol. 47, no. 5, pp. 425–429, 2009.

[7] ——, "A regional poison information center IVR medication identification system: does it accomplish its goal?" *Clinical Toxicology*, vol. 49, no. 9, pp. 858–861, 2011.

[8] J. A. Cramer, A. Roy, A. Burrell, C. J. Fairchild, M. J. Fuldeore, D. A. Ollendorf, and P. K. Wong, "Medication compliance and persistence: terminology and definitions," *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, vol. 11, no. 1, pp. 44–47, 2008.

[9] L. Osterberg and T. Blaschke, "Adherence to medication," *New England Journal of Medicine*, vol. 353, no. 5, pp. 487–497, 2005.

[10] A. S. Kesselheim, A. S. Misono, W. H. Shrank, J. A. Greene, M. Doherty, J. Avorn, and N. K. Choudhry, "Variations in pill appearance of antiepileptic drugs and the risk of nonadherence," *JAMA Internal Medicine*, vol. 173, no. 3, pp. 202–208, 2013.

[11] A. S. Kesselheim, K. Bykov, J. Avorn, A. Tong, M. Doherty, and N. K. Choudhry, "Burden of changes in pill appearance for patients receiving generic cardiovascular medications after myocardial infarction: cohort and nested case-control studies," *Annals of Internal Medicine*, vol. 161, no. 2, pp. 96–103, 2014.

[12] B. L. G. Van Wijk, O. H. Klungel, E. R. Heerdink, and A. de Boer, "Generic substitution of antihypertensive drugs: does it affect adherence?" *The Annals of Pharmacotherapy*, vol. 40, no. 1, pp. 15–20, 2006.

[13] M. Ude, K. Schuessel, R. Quinzler, K. Leuner, W. E. Müller, and M. Schulz, "Generic switch after ramipril patent expiry is not associated with decreased pharmacy refill compliance: a retrospective study using the DAPI database," *Journal of Hypertension*, vol. 29, no. 9, pp. 1837–1845, 2011.

[14] J. D. Stein, N. Shekhawat, N. Talwar, and R. Balkrishnan, "Impact of the introduction of generic latanoprost on glaucoma medication adherence," *Ophthalmology*, vol. 122, no. 4, pp. 738–747, 2015.

[15] J. Rathe, M. Andersen, D. E. Jarbøl, R. dePont Christensen, J. Hallas, and J. Søndergaard, "Generic switching and non-persistence among medicine users: a combined population-based questionnaire and register study," *PloS One*, vol. 10, no. 3, p. e0119688, 2015.

[16] C. Hall, "Technology for combating counterfeit medicine," *Pathogens and Global Health*, vol. 106, no. 2, pp. 73–76, 2012.

[17] DailyMed. (Sept. 12 2016). [Online]. Available: https://dailymed.nlm.nih.gov/dailymed/index.cfm

[18] Pillbox - National Library of Medicine. (Sept. 9 2016). [Online]. Available: https://pillbox.nlm.nih.gov/pillimage/search.php

[19] Pill Identification - WebMD: Identify Drugs by Picture, Shape, Color or Imprint - Pill Identifier with Pictures. (Sept. 9 2016). [Online]. Available: http://www.webmd.com/pill-identification/

[20] Pill Identifier (Pill Finder Wizard) - Pill Identification Tool Using Drug Pictures, Color, Shape, Number, or Imprint. (Sept. 9 2016). [Online]. Available: http://www.rxlist.com/pill-identification-tool/article.htm

[21] Pill Identifier - See What Your Drugs Look Like - GoodRx. (Setp. 15 2016). [Online]. Available: http://www.goodrx.com/pill-identifier

[22] J. J. Caban, A. Rosebrock, and T. S. Yoo, "Automatic identification of prescription drugs using shape distribution models," in *2012 19th IEEE International Conference on Image Processing*, 2012, pp. 1005–1008.

[23] Y.-B. Lee, U. Park, A. K. Jain, and S.-W. Lee, "Pill-ID: Matching and retrieval of drug pill images," *Pattern Recognition Letters*, vol. 33, no. 7, pp. 904–910, 2012.

[24] R.-C. Chen, Y.-K. Chan, Y.-H. Chen, and C.-T. Bao, "An automatic drug image identification system based on multiple image features and dynamic weights," *International Journal of Innovative Computing, Information and Control*, vol. 8, no. 5(A), pp. 2995–3013, 2012.

[25] A. Cunha, T. Adão, and P. Trigueiros, "Helpmepills: A mobile pill recognition tool for elderly persons," *Procedia Technology*, vol. 16, pp. 1523–1532, 2014.

[26] J. Yu, Z. Chen, S.-I. Kamata, and J. Yang, "Accurate system for automatic pill recognition using imprint information," *IET Image Processing*, vol. 9, no. 12, pp. 1039–1047, 2015.

[27] C. D. Manning, P. Raghavan, and H. Schtze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[28] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[29] RxIMAGE API - U.S. National Library of Medicine. (Sept. 23 2016). [Online]. Available: https://lhncbc.nlm.nih.gov/rximage-api

[30] RxNorm API. (Sept. 23 2016). [Online]. Available: https://mor.nlm.nih.gov/download/rxnav/RxNormAPIs.html#

[31] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv:1405.3531*, 2014.

[32] R. Bonney, C. B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. V. Rosenberg, and J. Shirk, "Citizen science: A developing tool for expanding science knowledge and scientific literacy," *BioScience*, vol. 59, no. 11, pp. 977–984, 2009.